16-833: Robot Localization and Mapping, Spring 2024

Homework 4

Dense SLAM with Point-based Fusion

Jinjiang You (jinjiany)

Due: Wednesday April 17, 11:59pm, 2024

2 Iterative Closest Point (ICP) (50 points)

2.1 Projective data association

u and v are the image pixel coordinates. They should satisfy:

$$0 \le u < W$$
$$0 \le v < H$$

d is the depth. It should satisfy:

$$0 \le d$$

We need to use $|p - q| < d_{thr}$ as a second filter because q may not be the nearest neighbour of p. q is p's projective nearest neighbour. The distance between these two 3D points can still be very large.

2.2 Linearization

Note that:

$$\delta R = \begin{bmatrix} 1 & -\gamma & \beta \\ \gamma & 1 & -\alpha \\ -\beta & \alpha & a \end{bmatrix} = I + \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}_\times$$

Therefore:

$$r_i = n_{q_i}^\top \left(\left(I + \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}_\times \right) p_i' + \delta t - q_i \right) \tag{1}$$

$$= n_{q_i}^\top p_i' + n_{q_i}^\top \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}_\times p_i' + n_{q_i}^\top \delta t - n_{q_i}^\top q_i \tag{2}$$

$$= n_{q_i}^\top p_i' + \left(p_i' \times n_{q_i} \right)^\top \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} + n_{q_i}^\top \delta t - n_{q_i}^\top q_i \tag{3}$$

Therefore:

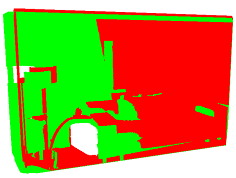
$$A_i = \begin{bmatrix} (p_i' \times n_{q_i})^\top & n_{q_i}^\top \end{bmatrix}$$
$$b_i = n_{q_i}^\top \left(p_i' - q_i\right)$$

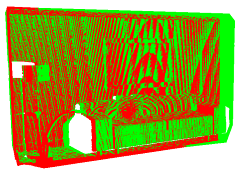
2.3 Optimization

We need to solve:

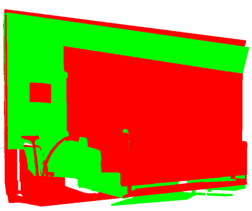
$$\left(\sum_{i=1}^{n} A_i^\top A_i\right) \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ t_x \\ t_y \\ t_z \end{bmatrix} = -\left(\sum_{i=1}^{n} A_i^\top b_i\right)$$

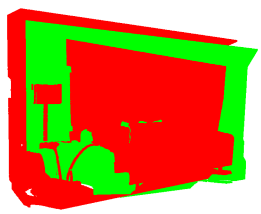
The results before and after ICP with frame 10 and 50 are shown below. The loss firstly increases because the number of inliers increases. Then the loss decreases to a very small number because the ICP successfully registers the two frames.





The results before and after ICP with frame 10 and 150 are shown below. The loss does not decrease and the number of inliers does not increase, which means the ICP fails. This is because the two frames differ too much. Compared with nearest neighbour, projective nearest neighbour method requires that the relative transform can not be too large. Otherwise, it will fail to find good correspondences between source and target points.





3 Point-based Fusion (40 points)

3.1 Filter

See code.

3.2 Merge

The weight average of the positions should be:

$$\frac{w \cdot p + 1 \cdot (R_c^w q + t_c^w)}{w+1}$$

The weight average of the normals should be:

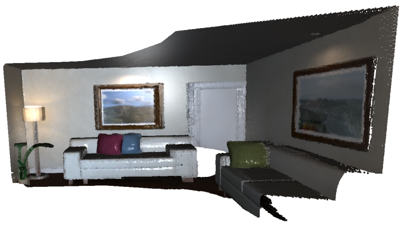
$$\text{normalized}\left(\frac{w \cdot n_p + 1 \cdot (R_c^w n_q)}{w+1}\right)$$

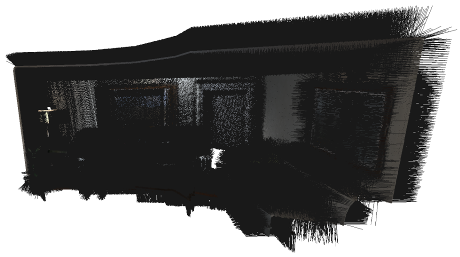
3.3 Addition

See code.

3.4 Results

The results of fusing 200 frames are shown below:

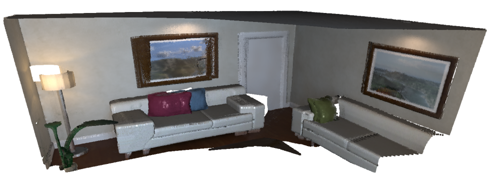


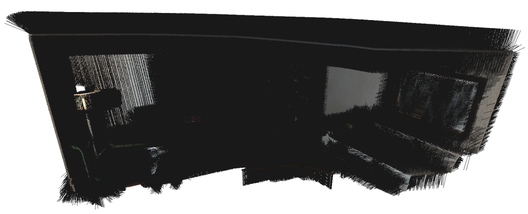


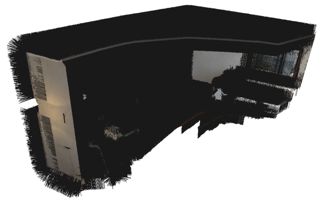
The total number of points in the fused map is 1362143. The number of points from the input frames is $200 \times (640 \times 480/4) = 15360000$. The compression rate is 8.868%.

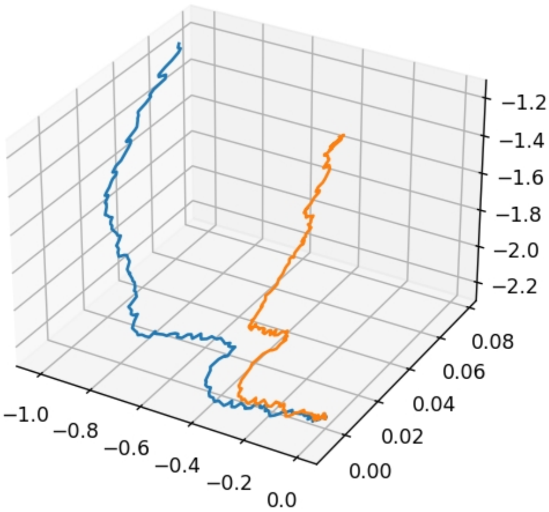
4 The dense SLAM system (20 points + 10 points)

In ICP, the source should be the map, and the target should be the input RGBD frames. We cannot swap their roles. This is because we use projective nearest neighbour to compute correspondences. The target points should be structured (i.e. have camera intrinsics parameters), so that the source can be projected into target's image space.







The visualization of the reconstructed scene looks good. However, according to the visualization of the trajectory, the error accumulates as the processing proceeds. This can be solved by frame-to-model tracking or a loop detection module.